

What is claimed is:

1. A data protection system for protecting files on a fileserver, the system comprising:

- a primary repository in communication with the fileserver via a network, the primary repository having:

- a primary repository node operative to store data;

- a primary repository node API in communication with the primary repository node and with the network and operative to communicate with the fileserver;

- a primary repository file transfer module in communication with the network and with the primary repository node and adapted for receiving files from the fileserver;

- a data mover in communication with the primary repository node API and operative to supervise the replication of files from the fileserver to the primary repository node;

- a location component in communication with the data mover and operative to store file location data;

- a directory service operative to maintain storage state for the primary repository node; and

- a node manager in communication with the location component and with the directory service and operative to manage primary repository node storage capacity and performance.

2. The data protection system of claim 1 wherein the system further comprises:

- a fileserver having:

- a filter driver operative to intercept input/output activity initiated by client file requests and to maintain a list of modified and created files since a prior backup;

a file system in communication with the filter driver and operative to store client files;

a policy cache operative to store a protection policy associated with a set of files;

a mirror service in communication with the filter driver and with the policy cache, the mirror service operative to prepare modified and created files in a share to be written to the primary repository node as specified in the protection policy associated with the set of files;

a location cache in communication with the mirror service and operative to indicate which repository should receive an updated version of an existing file; and

a location manager coupled to the location cache and operative to update the location cache when the system writes a new file to a specific repository node.

3. The system of claim 2 wherein the mirror service directs new versions of an existing file to the repository to which prior versions of the file were written.

4. The system of claim 2 wherein the system further comprises:

a fileserver API coupled to the mirror service and operative to communicate with a repository; and

a fileserver file transfer module in communication with the file system and operative to transfer files from the file system to at least one repository.

5. The system of claim 4 wherein the primary repository further comprises:

a protection policy component in communication with the data mover and operative to determine whether new versions of existing files should be compressed and whether older versions of existing files should be maintained.

6. The system of claim 5 wherein the system further comprises:

a remote repository in communication with the primary repository via a network, the remote repository having:

a remote repository node operative to store data;

a remote repository node API adapted for communicating with the remote repository node and with the network;

a remote repository file transfer module in communication with the primary repository file transfer module and with the remote repository node and adapted for receiving files from the primary repository file transfer module;

a data mover in communication with the remote repository API and operative to supervise the replication of files from the primary repository node to the remote repository node; and

a location component in communication with the data mover and operative to store file location data;

a directory service operative to maintain storage state for the remote repository node; and

a node manager in communication with the location component and with the directory service and operative to manage primary repository node storage capacity and performance.

7. The system of claim 2 wherein the protection cache is operative to define which repositories are used, how often data protection occurs,

how many replicas are maintained within each repository, and how modifications to share data are maintained.

8. A method for managing node managers in a repository having a plurality of nodes with associated node managers, the method comprising:

- starting the node managers in a bootstrap state;
- selecting a master node manager and a replica node manager according to specified criteria;
- setting all remaining node managers to drone state; and
- if at least one of the master and replica node managers fails, then selecting a replacement node manager from the drone node managers according to the specified criteria.

9. The method of claim 8 wherein selecting a master node manager and a replica node manager according to specified criteria comprises:

- determining a repository node with the lowest IP address and selecting the node manager associated with that repository node as the master node manager; and

- determining a repository node with the next lowest IP address and selecting the node manager associated with that repository node as the replica node manager.

10. The method of claim 8 wherein if the master node manager fails, then the method replaces the master node manager with the replica node manager.

11. The method of claim 8 wherein if the replica node manager fails, then the method replaces the replica node manager by determining a

repository node with the next lowest IP address and selecting the node manager associated with that repository node as the replica node manager